

Loomuliku keele töötlemine (NLP),  
tehisintellekt  
ning tavamõtlemine  
(common sense reasoning)

# Loengu teemad

- NLP tüüpilised ülesanded
- Vektorsemantika ja masinõpe
- Loogikapõhine semantika
- Common sense reasoning

NLP tüüpilised ülesanded

# • Kinds of NLP processing goals

Highly different methods for NLP, depending on the goals we have:

- Text similarity detection
- Automatic summarization
- Machine translation
- **Sentiment analysis**
- Speech processing
- Summary generation
- Dialog maintenance
- **Information extraction**
- **Question answering**
- ...

Meeldetuletus:  
vestlev pseudointellekt  
ELIZA

# Eliza: seletustega näitevestlus

ELIZA: HELLO, I'M ELIZABETH. WHAT WOULD YOU LIKE TO TALK ABOUT?

*(the Welcome message)*

**USER: My mum is having a hard time.**

ELIZA: TELL ME MORE ABOUT YOUR FAMILY

*(line 5 changes 'mum' into 'mother', and this is matched at line 26; hence one of lines 28, 29, and 30 is randomly selected)*

**USER: There's just her, dad and my brother**

ELIZA: WHAT DO YOU REMEMBER MOST ABOUT YOUR CHILDHOOD?

*(line 6 changes 'dad' into 'father', and this is matched at line 27; hence either line 29 or 30 is randomly selected for display - line 28 is ruled out because it was the last response to be selected from this group)*

**USER: My father and brother were always quarrelling.**

ELIZA: ARE YOU THE YOUNGEST IN YOUR FAMILY?

*('father' is matched at line 27; hence either line 28 or 29 is randomly selected for display - line 30 is ruled out because it was the last response to be selected from this group)*

**USER: My brother is younger than me.**

ELIZA: SO YOU ARE OLDER THAN YOUR BROTHER.

*(line 31 matches, so line 32 then generates the response 'SO me IS OLDER THAN my brother'. Output substitutions then transform this in turn to 'SO YOU IS OLDER THAN my brother' (line 10), 'SO YOU IS OLDER THAN YOUR brother' (line 13), and finally 'SO YOU ARE OLDER THAN YOUR brother' (line 18).*

Meeldetuletus:  
küsimustele vastamise süsteem  
IBM Watson

# IBM Watson võitis Jeopardy aastal 2011





- It's just a bloody nose! You don't have this hereditary disorder once endemic to European royalty : *Haemophilia*
- You just need a nap! You don't have this sleep disorder that can make sufferers nod off while standing up : *Narcolepsy*
- Heitor Villa-Lobos dedicated his "12 Etudes" for this instrument to Andres Segovia : *Guitar*
- Paganini's "24 Capricci" set the standard for etudes for this instrument : *Violin*
- Rembrandt's Biblical Scene "Storm on the Sea of" this was stolen from a Boston museum in 1990 : *Galilee*

# Masinõpe ja NLP

# • Example: sentiment analysis

Example letter from a customer:

“Dear <hardware store>

~ Yesterday I had occasion to visit <your competitor>. The had an excellent  
~ selection, friendly and helpful salespeople, and the lowest prices in town.

~ You guys suck.

~ Sincerely,”

## Teksti jätkumist ennustavad süsteemid:

- \* Bert (Google) ja selle derivaadid
- \* GPT-2, GPT-3 (OpenAI)

Mida ja kuidas nad teevad:

- \* Suure hulga tekstide pealt vektorsemantika jaoks statistikat:

mis sõnad/sõnaosad järgnevad etteantud tekstile.

- \* Kui ei leia olemasolevate seast täpselt etteantud teksti, hindavad

vektorsemantika järgi sarnasust olemasolevatele.

- \* Pakuvad selliselt üldistatud statistika pealt järgmist sõna. Siis järgmist.

Siis järgmist jne.

Mis on puhtalt masinõppe/statistikapõhise  
NLP meetoditega raske teha?

*Masinõpe ja inimkeel: praegused tipp-süsteemid ei saa hakkama selliste ülesannetega:*

- Linnavalitsus ei andnud meeleavaldajatele kogunemiseks luba, sest nad kartsid vägivalda. Kes kartis vägivalda?
- Palgil on kuus konna. Kaks lähevad ära ja kolm uut tulevad kohale. Mitu konna on nüüd palgil?
- Ma viisin eile oma riided pesumajja ja pean neile varsti järele minema. Kus mu riided praegu on?

A transcript of GPT3 <https://beta.openai.com/> playground

Q: Can stones drive a car?

A: No, stones cannot drive a car.

Q: Can wheels drive a car?

A: Yes, wheels can drive a car.

Q: Can water drink?

A: No, water cannot drink.

Q: Can beer drink?

A: Yes, beer can drink.

A transcript of GPT3 <https://beta.openai.com/> playground

Q: How fast does a car go from Washington to New York?

A: A car goes from Washington to New York in 4 hours.

Q: How fast does a bicycle go from Washington to New York?

A: A bicycle goes from Washington to New York in 2 hours.

Q: Three persons entered an empty room. One person left. How many persons are in the room now?

A: Two

Q: 10 persons entered an empty room. Two more came. Five left. How many persons are in the room now?

A: 10



# Klassikaline loogika ja NLP

## *Näiliselt tore, mikro / mängunäited töötavad:*

Isa(Jaan,Mihkel).

Isa(Jaan,Ants).

Isa(Ants,Peeter).

Iga  $x, y, z$  jaoks:  $\text{Isa}(x,y) \ \& \ \text{Isa}(y,z) \Rightarrow \text{Vanaisa}(x,z)$ .

Tõesta, et eksisteerivad  $z, u$  nii et  $\text{Vanaisa}(z,u)$ .

**NB!** Selles näites ei ole sõnade „Isa”, „Jaan” jne jaoks loogika mõttes teada tegelikult mitte midagi, peale nendesamade faktide / reeglite, mis siin kirjas.

## *Klassikalist loogikat on NLPs lootusetu kasutada:*

Loomulikust keelest arusaamine eeldab nõ **tavamõistust (common sense) teadmisi + arusaama/järeldusoskust**

Klassikaline loogika ei sobi, sest:

- \* Vastuolu teadmibaasis ei ole OK: sellest saab järeldada midaiganes
- \* Ebakindlaid teadmisi ei saa olla
- \* Vaikimisi eeldusi (tüüpiliselt ...) ei saa olla
- \* Sarnasustest ei saa midagi järeldada
- \* Olukorrad / faktid ei saa ajas muutuda
- \* Kõik teadmised tuleb eksplitsiitselt kirja panna
- \* ... jne ...

Hübriidsüsteemide poole

## *Liites masinõppe / vektorsemantika ja loogikapõhise semantika, võiksime saada parema süsteemi*

- Levinud suund: püüame vektorsemantika / masinõppe süsteemidele reeglivärki juurde ehitada.
- Meie grupi suund: modifitseerime reeglisüsteemide loogikat „pehmeks” ja püüame sinna vektorsemantikat / masinõppe värke juurde ehitada

# Reasoning with uncertainty

# Uncertainties everywhere

Almost all the knowledge we have is uncertain: there are many **exceptions** to a rule or a fact/rule holds with some **vague probability**.

Notice that, in contrast, typical databases in companies contain facts with very high certainty, but they do not contain rule or commonsense knowledge: the latter is built into brittle software code of systems using these databases.

In contrast, mathematics is certain, **because we define it**

# Discrete and numeric methods

In practice there are two main different ways of tackling uncertainty:

- **Numeric methods**: uncertainty is estimated with probability-like numbers of various kinds.
- **Discrete methods**: no numbers used. Instead, uncertainty is described as exceptions or beliefs of people.



# Discrete example from default logic: birds

bird(tweety).

penguin(pennie).

penguin(X)  $\Rightarrow$  bird(X).

penguin(X)  $\Rightarrow$  -flies(X).

Use (bird(X)  $\Rightarrow$  flies(X)) if we cannot prove that  $\neg$ flies(X).

Query: flies(X) ?

Common syntax for  
default rules:

$$\frac{\text{bird}(X) : \text{flies}(X)}{\text{flies}(X)}$$

# Another classic example from default logic

republican(nixon).

quaker(nixon).

Use  $(\text{republican}(X) \Rightarrow \neg \text{pacifist}(X))$  if we cannot prove that  $\text{pacifist}(X)$ .

Use  $(\text{quaker}(X) \Rightarrow \text{pacifist}(X))$  if we cannot prove that  $\neg \text{pacifist}(X)$ .

Query:  $\text{pacifist}(\text{nixon})$  ?

# Some notes about default logic

- Similar to the somewhat vague concept of “defeasible logic”
- Since first order logic (FOL) is undecidable, it is generally impossible to check that a rule can be fired!
- Practical systems are almost all focused on propositional instances of FOL formulas: propositional logic is decidable, thus rules can be checked.
- Nonmonotonic logic: all standard logics are monotonic, default logic is not

# Numeric methods: an example

John observes something in the field which looks like a bird and estimates the probability of it being a bird as 80%. Mike observes the same object, but estimates the probability as 70%. I have an intuitive rule which says that birds can fly with the probability 90%. Let us formalize this as:

$\text{bird}(\text{object1}) : 0.8 :: \text{John}$

$\text{bird}(\text{object1}) : 0.7 :: \text{Mike}$

$[\text{bird}(X) \Rightarrow \text{canfly}(X)] : 0.9 :: \text{me}$

What can I derive from here? A simple idea is to combine the bird observation probabilities to a stronger one, using a standard probability calculation rule  $P(a \vee b) = P(a) + P(b) - (P(a) * P(b))$ , which holds in case a and b are independent observations. We get:

$\text{bird}(\text{object1}) : 0.8+0.7-(0.9*0.7) = 0.87 :: \text{John, Mike}$

Second, using the derivation rule, decreasing the probability according to a probability rule  $P(a \ \& \ b) = P(a) * P(b)$ , again in case when a and b are independent:

$\text{canfly}(\text{object1}) : 0.87 * 0.9 = 0.783 :: \text{John, Mike, me}$

# What can go wrong in this example?

We have to be careful to indicate what the probability number means: it can be interpreted in several ways. For example, does

`bird(object1) : 0.1`

mean that object1 is unlikely to be a bird, or that it is a bit likely to be a bird? In other words, are small probability numbers really interpreted as positive or rather indications of a probability of a negation

`-bird(object1) : 0.9`

What more could go wrong or require extra care?

- Are the calculation rules we used really true or applicable? In which context are they applicable? Can they lead to incorrect or unintuitive results in other cases?
- What about "independent": are these statements independent and to what degree?
- Are we sure our procedure does not use same input several times to "over-strengthen" the probabilities?
- What is the "formalism" we used, ie how to encode that a rule/fact has a probability and sources? Can we use quantifiers also there, like quantifying over sources or something?

# Different types of uncertainties

There are many different camps of people advocating different ways to understand and handle uncertainty. The main camps are:

- **Frequentists**: probabilities as limit values of statistical experiments, when more and more experiments are made. Dice-kinds of examples fit this camp very well. Also called **physical** interpretation.
- **Bayesianists**: more subjective, interpreting probability as "reasonable expectation" (a probability is assigned to a hypothesis), yet basing their calculations on the famous [Bayes theorem](#). Also called **evidential** interp.

Bayesianists are similar to

- **Subjectivists**: probability measures a "personal belief" (think making bets)

# Different measures of uncertainty

Broadly said, the main two measures are:

- **Standard**:  $\text{red}(a1): p$  indicates that the probability of  $a1$  being red is  $p$ .
- **Fuzzy**:  $\text{red}(a1): p$  indicates how red  $a1$  is: (1.0 means totally red)

Fuzzy measures lead to **fuzzy logic**, which is a popular field.

# Intervals of probability

It may happen that we know the **lower and upper limits of probability**, but not the exact probability.

Fits well for modeling sensors which can sometimes give erroneous values, while we know – from stats – their error rates.

The most famous approach in this direction is the **Dempster-Schafer theory of evidence**.



Kontseptsioonid

faktid

situatsioonid

**On top level, there are three main types of assertion sentences:**

- \* **Conceptual statements** like "Cats are animals", "Animals have two legs", which do not describe a specific situation and are not dependent on uncommon circumstances. Typically they describe concepts.
- \* **Fact statements** like "Mozart was a composer", "Tallinn is a capital of Estonia", which describe named entities, again, not dependent on uncommon circumstances.
- \* **Situational statements**, which describe a concrete situation and events happening with/in this situation, like "A cat was catching mice", "Mozart sat down and started playing a piano".

## Scenes, situations and events

We structure scenes into situation-event chains with the common form

situation1 - event1 - situation2 - event2 - situation3 - ...

following this principle:

all properties of the next situation are assumed to be the same as those of the previous situation, except when we specifically/with high confidence know that they are not.

This principle can be formalized using default rules (normally) or frame axioms (in simple situations). Example of a chain of events and situations follows.

"A man entered a room containing a table. He wore a black suite. Then the man left the room".

Näiteid reaalse süsteemiga

# Näiteid

- \* Suurem pingiviini-näide
- \* Klassifitseerimise näide
- \* Social network näited
- \* Sarnasuse näide
- \* Situatsiooni näide
- \* quasimodo knowledgebase näited